

Themantic Education's

IB Theory of Knowledge

A Student's Guide

eBook

FREE



UNIT 3

**Bias in personal
knowledge**

Alexey Popov



UNIT 3 - Bias in personal knowledge

Exhibition: a turbulence map	72
Story: Senate Bill 464	73
Lesson 1 - Bias	74
Lesson 2 - Personal experience	77
Lesson 3 - Darwinian evolution of personal knowledge	81
Lesson 4 - Analogy analysis	85
Lesson 5 - Cultural experience	89
Lesson 6 - Memes and Universal Darwinism	93
Lesson 7 - Heuristics	97
Lesson 8 - Implicit bias and bias self-awareness	101
Lesson 9 - Bias reduction	105
Lesson 10 - Compos mentis	109
Back to the exhibition	113

UNIT 3 - Bias in personal knowledge

You may remember from Unit 1 (“Knowledge of knowledge”) that there is a distinction between **personal knowledge** and **shared knowledge**. These terms are quite transparent: personal knowledge is something belonging to you as an individual, while shared knowledge is something common to sizeable groups. Shared knowledge and personal knowledge are overlapping circles on a Venn diagram. Some of your personal knowledge coincides with that shared by other people, but another part of your personal knowledge is unique to you.

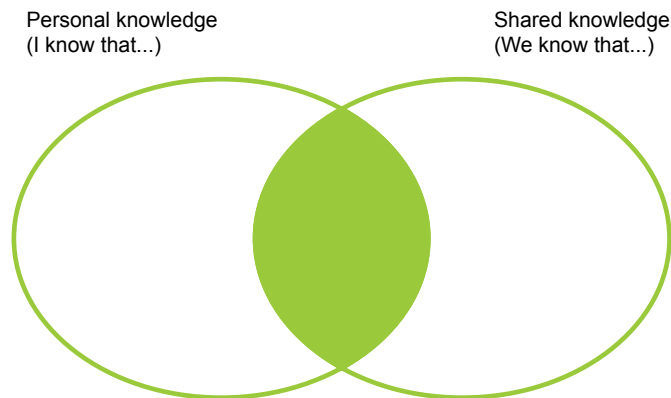


Image 1. Personal knowledge and shared knowledge: how they are related

The good thing about bias is that, although every individual is biased, collectively we can keep these biases in check and overcome them. In a series of independent replications, conclusions of one scientist may be validated by other scientists. In a jury court, opinions of the jurors may be compared and discussed. Scientists may have different explanations for an observed phenomenon, but through testing and replication some explanations are eliminated and some retained. In other words, biases are abundant in the realm of personal knowledge, but not so much in shared knowledge. As a rule, shared knowledge is much less biased than personal knowledge.

Can biased personal opinions be valuable for developing shared knowledge? (#Perspectives)

The bad news is that shared knowledge can also be biased. Biased shared knowledge is probably more disastrous than biased personal knowledge simply because we trust it more. Additionally, it is much more difficult to identify the bias and eliminate it when it is the whole of humanity that is biased. In other words, although biases in shared knowledge are less numerous, they are more impactful.

	How many biases are there?	How impactful are they?
Personal knowledge	A lot!	They affect only you
Shared knowledge	Not so many	They affect everyone!

KEY IDEA: Biases in shared knowledge are less numerous, but they are more impactful

In this unit, we will consider biases in personal knowledge. On the surface, the problem may seem simple: just check your personal knowledge against shared knowledge and get rid of your bias! However, we cannot just dismiss personal knowledge as something inferior to shared knowledge. After all, as a knower, your personal knowledge is all you have access to. A belief that you retrieve from your personal knowledge can either come from the area that overlaps with shared knowledge or from the area that is uniquely yours. How do you know which area it comes from?

Can we know if our personal knowledge is biased without checking it against shared knowledge?
(#Methods and tools)

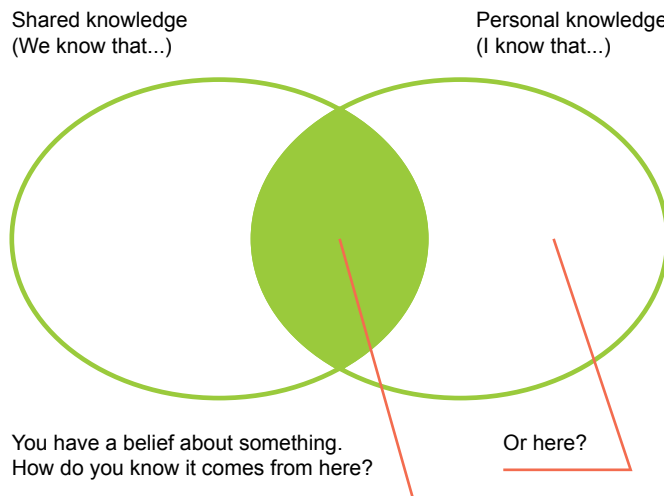


Image 2. Where does your belief come from?

The knowledge that you are directly in touch with and that you use on a daily basis is your personal knowledge. For this reason, personal knowledge is worth considering on its own before we move on to biases in shared knowledge.

Exhibition: a turbulence map

In front of me is an aviation weather forecast chart (for simplicity I will call it a turbulence map).

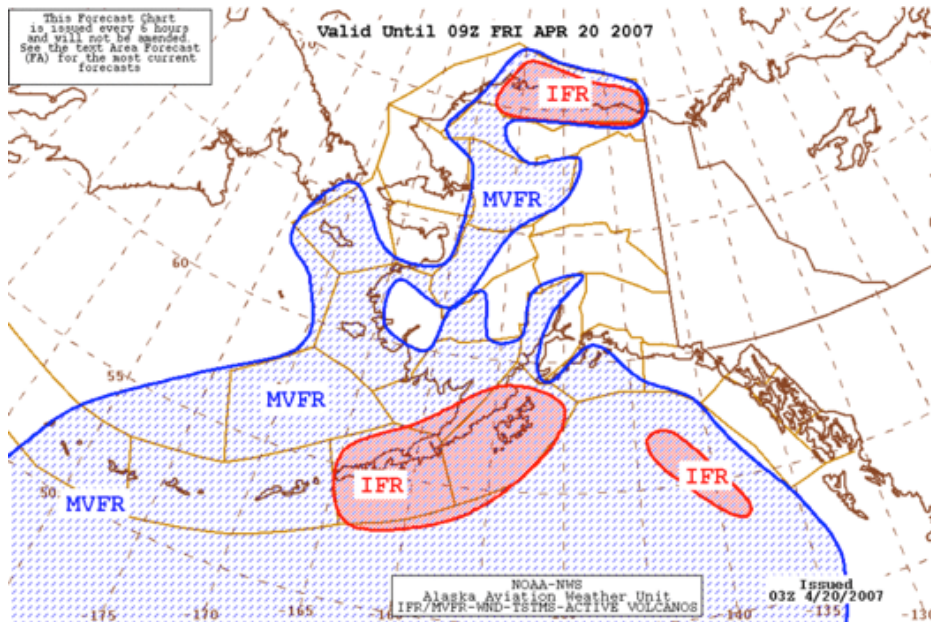


Image 3. Aviation weather forecast chart (turbulence map) (credit: Wikimedia Commons)

Such maps show you the areas where turbulence is more likely to occur when you are travelling by air. These maps (among other sources of information) are used by pilots to try to make your flight smoother when they are navigating.

I am a nervous flyer. I have a complicated relationship with turbulence. It is pretty unfortunate for someone who works in an international setting and needs to travel a lot.

In front of me is an aviation weather forecast chart (for simplicity I will call it a turbulence map).

Such maps show you the areas where turbulence is more likely to occur when you are travelling by air. These maps (among other sources of information) are used by pilots to try to make your flight smoother when they are navigating.

I am a nervous flyer. I have a complicated relationship with turbulence. It is pretty unfortunate for someone who works in an international setting and needs to travel a lot.

At some point when it became really irritating, I started educating myself. I read articles and watched videos that explained turbulence and analyzed past airplane crashes. I discovered that a lot of my beliefs had been inaccurate and misleading. First of all, I used to think that turbulence can cause airplanes to crash. Now I know that airplanes are designed so that they can withstand turbulence more than two times stronger than anything commercial flights are likely to encounter. I used to think turbulence was the most dangerous part of the flight. Now I know that you are more likely to be harmed while you are on the tarmac than when you are experiencing turbulence mid-air. I used to think air travel was a risky option. Now I know that statistically I am much more likely to die in a car on the way to the airport.

Has it helped? No. Every time turbulence kicks in, I still grab the armrest until my knuckles turn white. In reality, I should be doing that in taxis, not in planes! My conscious brain knows that, but my body seems to refuse to listen.

I still check “turbulence maps” before flying. The abundance and accessibility of such maps online gives me a hint that I am not alone. It appears as though there are many more nervous flyers out there who misinterpret the danger of planes (relative to other means of travel), whose logical brain cannot override the rest of their brain, whose expectations, perceptions and attitudes to air travel are all biased because of this complicated relationship with turbulence.

The truth is, if your seatbelt is fastened, turbulence is not dangerous. My beliefs and perceptions, however, systematically deviate from this truth in the direction of misinterpreting various aspects of air travel as more dangerous than they really are.

Patrick Smith, an airline pilot and author, gives exhaustive answers to some common questions about air travel on his website, www.askthepilot.com

If you are interested in learning more about turbulence, there is a section on the website dedicated to that under “questions and answers”.

Story: Senate Bill 464

The year 2019 in the USA saw an unusual precedent in legislation: Senate Bill 464 made it mandatory for doctors and nurses in California to undergo eight hours of implicit bias training and testing periodically (every 2 years).

This is probably one of the first times when the concept of implicit (unconscious) biases made its way into legislature.

This bill was “inspired” by some disturbing research findings that showed that, although there was a decrease in the overall number of women who died giving birth in California, black women were still 3 or 4 times more likely to die from complications at childbirth compared to white women. Additional research into this issue showed that roughly half of surveyed medical professionals believed myths and shared misconceptions about racial differences in tolerating pain. For example, they believed that black patients can “endure more pain” and have “thicker skin”. Such biases created a situation where, when an expectant black mother claimed she was in pain, doctors underestimated the severity of her condition and did not respond appropriately. Obviously, the medical professionals were entirely oblivious of this bias that they had. This research was conducted in 2016. While it is quite hard to believe that such racial biases are so widespread in the 21st century, we cannot simply attribute this to “bad doctors”. These biases are implicit – they occur without the conscious awareness.

The bill requires medical professionals to go through training that teaches them to identify their own implicit biases and consciously counteract them. This is an attempt to reduce discrimination by targeting our own unconscious minds.

You can read more about the Bill in the article “These California bills would train nurses, judges and police how to spot their own biases” in *Los Angeles Times*.



Image 4. There are racial differences in the chance of death from complications at childbirth

The full text of the bill can also be found online, its name is SB-464, California Dignity in Pregnancy and Childbirth Act.



Lesson 1 - Bias

Learning outcomes

- [Knowledge and comprehension] What is bias?
- [Understanding and application] What are the key examples and non-examples of bias?
- [Thinking in the abstract] How can bias be separated from similar knowledge concepts (such as prejudice, misconception or superstition)?

Plan

In this lesson we will define bias and think about examples and non-examples of bias. In line with the purpose of this unit, the focus will be on bias in personal knowledge.

Just to remind you, bias in personal knowledge may be assessed against shared knowledge. If we want to know if our personal belief is biased or not, we can compare it to the accepted, well-established beliefs on the same subject matter that we have collectively agreed upon.

Shared knowledge, of course, can also be biased, but that will be the focus of the next unit.

What is bias?

Is it true that we are much more biased than we could possibly imagine? (#Scope)

As much as I would like to think of myself as an open-minded, unprejudiced, impartial and just individual, I know that I am not one (are you?). Growing up, I was influenced by a variety of factors and exposed to a variety of experiences. In all probability, these experiences have caused me to have certain biased beliefs. Worst of all, I am probably biased in ways that I am not even aware of.

I will define **bias** as a *systematic deviation from the truth*.

When I say “*deviation*”, I imply that there exists a correct answer (belief, decision) and that the answer (belief, decision) we are dealing with does not match this correct one. This is important because we can identify a bias only if we know the correct answer. If we do not know what the correct answer is, or if we cannot at least assume the correct answer beyond a reasonable doubt, there is no point in talking about bias.

When I say “*systematic*”, I mean a deviation that is not random. In other words, it is leaning consistently towards one direction rather than various directions at various times. For example, suppose you are measuring the width of your bed with a measuring tape. You carry out the measurement 10 times. Every time you will get slightly different readings, both higher and lower than the real width of your bed. This is an example of *measurement error*, but this is not a *bias*. A bias occurs when, for some reason, the measurement deviates systematically in one direction. For example, suppose the measuring tape itself is flawed – you washed it accidentally in the washing

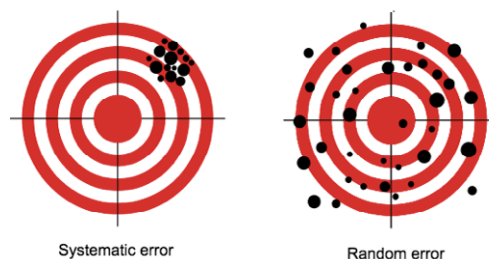


Image 5. The difference between systematic error and random error (credit: Wikimedia Commons)

Key concepts

Bias, systematic deviation, opinion, perspective, mistake

Other concepts used

Stereotype, prejudice, misconception, superstition, decision-making

Themes and areas of knowledge

Theme: Knowledge and the knower

machine and it shrank a little, resulting in each inch section being a little shorter than it is supposed to be (I am now assuming that it is a cloth measuring tape, not a metal one... why would you put a metal measuring tape in a washing machine?). In this case, no matter how many times you carry out the measurement, you will always underestimate the width of your bed. This is bias.

KEY IDEA: Bias is a systematic deviation from the truth

Sources of bias

Since the deviation is systematic, it is usually the case that the deviation is *caused* by something, in other words, that there is a **source of bias**. In my turbulence example, overestimating the dangers of air travel is caused by my fear of turbulence. It also probably means that whenever there is bias, we can identify one or several factors that make it happen.

Theoretically:

- If we can eliminate the source, the bias will disappear
- If we know the source, we can predict the bias (for example, knowing that a person has a fear of turbulence means that we can probably predict that they will overestimate the dangers of air travel)

Is there any way to know what causes our personal bias?
(#Methods and tools)

There are many possible sources of personal bias. Some of them are linked to our identity (cultural, political, gender). Some are linked to our personal experiences (having survived through certain difficulties, having witnessed certain events). Arguably, every human being has a different background and that could determine how (in what way) they are biased.

The important take-away message here is that biases are systematic because they are systematically affected by a certain source and, at least theoretically, these sources can be identified and dealt with.

Bias versus other concepts

To understand a concept, it is always useful to separate it from (misleadingly) similar concepts by answering the question “What is it not?”

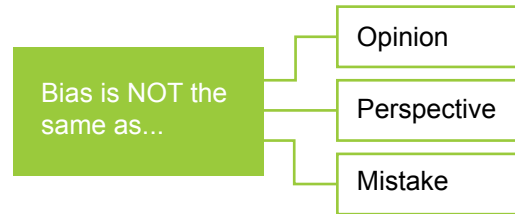
We have defined bias by stating what it is. Let us now try to delineate it from a variety of other concepts that it can be easily confused with.

Bias is not the same as **opinion**. Opinions are possible when there is no single truth. For example, it is my opinion that restaurant A is better than restaurant B. Airplanes falling because of turbulence cannot be my opinion because we do know that this is false. Since we have access to a pretty unambiguous truth in this case, opinions are no longer a thing – there are either beliefs that correspond to the truth or ones that don't.

Bias is not the same as **perspective**. Again, perspectives are possible when the truth is complex and when multiple interpretations of the truth are possible. For example, there may be various historical perspectives on events of the past. There can be various angles of looking at those events, and often there is no way to prefer one perspective over another. For this reason, perspectives are very valuable (the more the better!). By contrast, in my turbulence example, the truth is pretty straightforward. Another difference is that, when you are presenting a perspective, you are presenting it honestly as one of several possible angles in looking at a situation. You acknowledge the existence of other angles. When you are biased, you are trying to pass your bias off as the truth (and you actually believe it to be the truth).

Is it possible for biases to be accepted as valuable perspectives?
(#Perspectives)

Bias is not the same as a **mistake**. It is a particular type of mistake – a systematic one. If I ask a child who has never travelled by air if turbulence can bring down airplanes, they may say yes. It would be a mistake but not a bias. If you ask someone like me (before they educated themselves with loads of articles and videos), they will say yes because they are afraid of turbulence. They will answer multiple other questions with similar mistakes – for example, they will overestimate the likelihood of turbulence occurring, the psychological effect it has on airline pilots, and the number of turbulence-related accidents in the past. All of their answers will be biased in the same direction, driven by one source - their underlying fear of turbulence.



Critical thinking extension

Now that we are clear with the definition of bias and with some of the things that bias is *not*, can we name some examples of phenomena that may be categorized as instances of bias in personal knowledge?

Here are some of the phenomena that we are going to consider further on in this unit:

- 1) Biased perception (for example, susceptibility to certain perceptual illusions)
- 2) Stereotypes
- 3) Prejudice
- 4) Biased decision-making (for example, selecting risky options when it is not logically warranted)
- 5) Misconceptions (biased understanding of certain ideas, not just a mistake but a systematically incorrect understanding driven by a false belief)
- 6) Superstitions (stubborn beliefs in supernatural influences despite counter-evidence)

Do you think all of these phenomena fit our definition of bias equally well? Would you add any other phenomena to the list?

To what extent can we claim that personal bias penetrates every aspect of our lives? (#Scope)

If you are interested...

When a meteorologist talks about bias, it is worth listening to (I would know, both of my parents have degrees in meteorology). J. Marshall Shepherd's TED talk "3 kinds of bias that shape your worldview" (2018) is a good place to start.

Take-away messages

Lesson 1. Bias is a systematic deviation from the truth. This definition implies two things: (1) there exists a certain standard that we may accept as the correct answer or the truth, (2) the deviation from this standard is not occasional and random, but systematic (consistent and always in the same direction). For this reason, opinions, perspectives and mistakes are all non-examples of bias. Since biases are systematic, it must be the case that they are (systematically) influenced by some factors. Such factors are known as sources of bias and they can originate from your personal experiences, your culture, your identity, and so on.

Lesson 2 - Personal experience

Learning outcomes

- [Knowledge and comprehension] What is a personal experience sample?
- [Understanding and application] How do personal beliefs depend on personal experiences?
- [Thinking in the abstract] To what extent can we claim that personal beliefs are inevitably biased?

Recap and plan

We have defined bias as a systematic deviation from the truth. We have considered some examples and non-examples of bias in personal knowledge, so now we know what it is. Now we can ask ourselves, where does it come from? Why is our personal knowledge biased in the first place?

In this lesson I start building my argument by making the point that personal beliefs are likely to be biased because they are based on personal experiences that are inevitably limited. In doing so, I will define the concepts of personal experience and experience sample.

KEY IDEA: *Personal beliefs are likely to be biased because they are based on personal experiences that are inevitably limited*

Personal experience and experience sample

Personal experience is the sum total of all instances of interaction of a person with various aspects of the world. This is a broad definition that includes any type of interaction, both practical and theoretical. If you have seen a zebra on a safari trip, you have some personal experience with zebras. If you read or watched a documentary about zebras, you also have experience with them. When I say “zebra”, you have a complex of associations firing up in your brain – that is your personal knowledge about zebras based on your interactions with various aspects of reality (books, documentaries, safari parks) somehow connected to them.

However, what comes to your mind when I say “Bony Giant Sengi” or “elephant shrew”? I bet your experience with this animal is very limited, maybe even to the extent where your mind is blank.

If you have never seen this animal, nor heard anything about it, you have no personal experience with it.

Personal experiences are limited because the world is so vast that it is unrealistic to expect anyone to experience all aspects of it within one lifetime. Coming back to elephant shrews, there is an estimated 8.7 million different species of animals on the face of the Earth. How many of those species do you have personal experience with, either practical or merely theoretical?



Image 6. *What do you know about Bony Giant Sengi, a.k.a. elephant shrew?*
(credit: Kim, Flickr)

Key concepts

Personal experience, experience sample

Other concepts used

Representativeness, sample and target population (in the analogy with human sciences), limited versus biased

Themes and areas of knowledge

Theme: Knowledge and the knower

Is personal experience always inevitably limited?
(#Scope)

To define the aspects of the world that a person has had experience with, we will also use another term – experience sample. If 8.7 million species are the “total” number of available experiences, then the several dozen species I know something about will comprise my **experience sample**. My experience sample is probably different from yours. Take any two people and their experience samples will overlap, but not coincide – this applies not only to knowledge of animal species, but to anything!

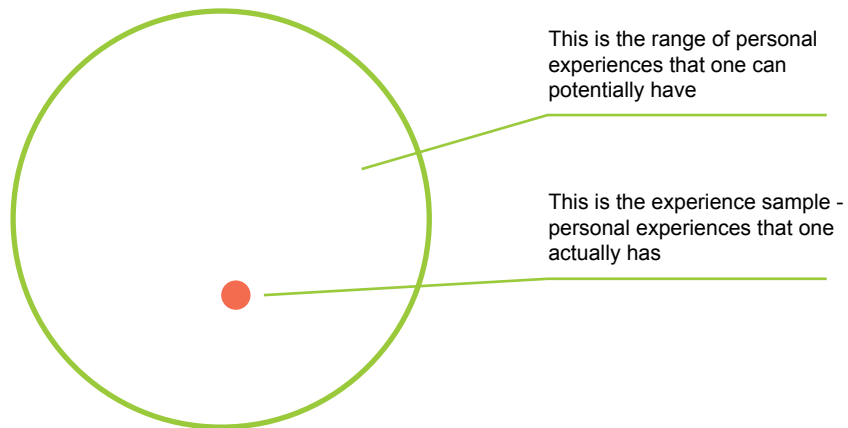


Image 7. Experience sample

Personal knowledge is the product of personal experience samples

To start, I will claim that our personal beliefs are based on our experience samples (which are very limited).

I describe myself as a devoted introvert. Knowing this, you will not be surprised to hear that one of my favorite pastimes is to look up the most remote places on the face of the Earth and dream about moving to live in one of these places one day.

To what extent is our personal knowledge the product of our personal experiences?
(#Perspectives)

Currently, the most remote settlement on Earth is a town with an exotic name, Edinburgh of the Seven Seas. The island upon which the settlement lies (Tristan da Cunha) has no airstrip, so the only way to travel there is by boat. The 2,810-kilometer boat ride from South Africa (the nearest location) takes around 6 days.

The settlement’s population is several hundred people. I cannot help but wonder, “What are they like? What would it be like to live there?” I know very little about those islanders, but I do have some beliefs. For example, I somehow find myself believing that inhabitants of Edinburgh of the Seven Seas are simple people who feel very attached to their home, but at the same time are extremely cautious about strangers.



Image 8. Tristan-da-Cunha: welcome to the remotest island



Image 9. Tristan-da-Cunha: aerial view

When I think about these beliefs (thinking about thinking, metacognition!), I realize that they are based on the very limited experiences that I have had. Namely, I remember reading somewhere that in 1961, there was a volcanic eruption on the island and the whole population had to abandon the settlement and was moved to the UK. Two years later, when it was declared safe again, they all chose to go back. I remember thinking, “Wow, these people like their home and don’t care about the gifts of civilization that we are all after.”

I have a personal belief about inhabitants of Edinburgh of the Seven Seas, and this belief is based on a very limited experience sample. Another person may have a slightly different experience sample, which would lead to a different belief.

If the experience sample is biased (which it is likely to be), then the personal knowledge based on it will also be biased.

We must have personal beliefs which are very likely to be biased

Once I arrived at this thought, my natural reaction was: “Well, you should make sure that your personal beliefs are not based on limited experience samples... gain more experience and only then form a personal belief!”

However, on reflection, it does not seem to be that simple, because:

- 1) Is it even possible to ever have *enough* personal experience with something to be *certain* that your personal belief is unbiased? Our personal experiences will always be limited. The world is too large for us to be able to experience every aspect of it. In fact, it looks like our personal experience is a tiny spotlight on a huge canvas that the world has to offer.
- 2) Once I accept that my personal experience is inevitably limited, can I opt out of having a personal belief at all? Rather than having a biased belief, I want to choose having no belief. But let’s face it, it does not seem to be possible. We need personal beliefs to navigate the world. They save us a ton of time and effort in a variety of everyday situations. Just think about it: when you are in a restaurant and a waiter approaches you, you do not expect the waiter to attack you, because you are operating on the assumption (expectation) that he is a decent person who is willing to serve you food. Imagine you did *not* have this belief about the waiter. He would have to gain your trust first, and that is a waste of his working hours.

Is it possible to have no belief at all rather than a biased belief?
(#Perspectives)

KEY IDEA: We have no other option but to have personal beliefs that are very likely to be biased

I am arriving at an interesting conclusion. Having personal beliefs is inevitable. We cannot not have personal beliefs. At the same time, personal beliefs are based on personal experiences, and personal experiences are (very) limited. We cannot ever have complete experience. This means that personal beliefs are likely to be biased. Hence, when it comes to our personal knowledge, having biased personal beliefs is a necessity that we cannot opt out of.

Well... isn’t this a little disappointing?

Critical thinking extension

The argument that I have been building in this lesson rests upon several key claims and assumptions:

- 1) We need personal beliefs in order to function in this world
- 2) Personal beliefs are based on personal experiences
- 3) Personal experiences are always limited

If you want to attack my argument (which you are welcome to do!), you probably need to target one or more of these statements. If any one of these statements is flawed, then the whole argument is flawed.

For example, you might want to attack the third claim. You might point out that “limited” personal experiences does not necessarily mean “biased”. We know from human sciences that a sample of participants is “limited” in relation to the population that the results will be applied to, but it is not “biased” if the sample is shown to be representative. In other words, if characteristics of the sample reflect the essential characteristics of the population, the sample is limited but not biased. Is the same logic possible with experience samples? This raises some interesting questions such as “How do you ensure that your personal experience sample is representative?”

Can personal experiences be representative of the world in the same way as samples in human sciences can be representative of the population? (#Methods and tools)

If you are interested...

In human sciences, if a sample is representative of the target population, it is believed not to be biased despite the fact that it is obviously limited. Representative samples allow researchers to apply the results from the sample to the whole population.

If you are not familiar with the concept of representativeness of a sample in human sciences or simply want to refresh this knowledge, you can watch the video “Selecting a representative sample” from the YouTube channel *Research By Design*. It discusses populations and samples and investigates how to make your sample representative of your population.

Take-away messages

Lesson 2. Personal knowledge is formed on the basis of personal experience. Personal experience samples are inevitably limited because the world is too large for someone to experience all aspects of it. Hence, personal knowledge will also inevitably be limited. To the extent that personal experience samples are biased (which they are likely to be), personal knowledge will also be biased. Moreover, we cannot opt out of having limited personal beliefs because we depend on these beliefs to navigate the world.

Lesson 3 - Darwinian evolution of personal knowledge

Learning outcomes

- a) [Knowledge and comprehension] What is Darwinian evolution?
- b) [Understanding and application] What are the similarities between the development of personal knowledge and adaptation of species through natural selection?
- c) [Thinking in the abstract] To what extent can we claim that development of personal knowledge is a Darwinian process?

Recap and plan

In the previous lessons I claimed that personal beliefs are inevitably based on personal experiences and that personal experiences are inevitably limited. Since personal beliefs are inevitably based on limited information, they are likely to be biased.

We have accepted the idea that personal experience somehow influences the development of personal knowledge. That being said, what exactly is the nature of this influence?

In this lesson I will suggest an **analogy** between the process of developing personal beliefs and the process of Darwinian evolution of species. My analogy will imply that the way personal experience influences personal knowledge is similar to the way the environment influences the process of natural selection.

Theory of evolution: quick refresher

Here is a quick refresher on the theory of evolution, as suggested by Charles Darwin (1809 - 1882) and modified slightly in more contemporary versions that followed the discoveries in genetics. These ideas are referred to as **Darwinian evolution**.

Key concepts

Darwinian evolution, analogy

Other concepts used

Environment, evolution, superstition, personal beliefs, natural variation, differential fitness, survival of the fittest, adaptation through natural selection, Universal Darwinism

Themes and areas of knowledge

Theme: Knowledge and the knower

What is the role of analogy in acquiring new knowledge?
(#Methods and tools)

What the theory claims	What it means
When two organisms have a baby (please excuse my French!), the baby's genotype is a random combination of the genotypes of the two parents. Because of this randomness, there is always some variation in the gene pool. This is called " <i>natural variation</i> ".	My child, welcome into this world. We will give you this genotype that we randomly created out of our own genes, and see what happens.
Survival of an organism depends on its fitness. Some organisms have genotypes that are more fit to the demands of the environment, some have genotypes that are less fit. This is known as " <i>differential fitness</i> ".	Your genotype will determine how well you fit into the environment.
Organisms that are more fit to the environment have higher chances of survival. This principle is known as " <i>survival of the fittest</i> ".	If you do not fit well, you will not be able to pass on your genes.
Through this process of survival of the fittest, generation after generation, genes that provide a good fit are more likely to stay in the gene pool while genes that provide a poor fit gradually disappear from the gene pool. This is known as " <i>adaptation through natural selection</i> ".	If you fit well, you will have children and pass your genes on to them.

How suitable is Darwin's evolutionary theory to explain historical development of knowledge? (#Scope)

Darwin's biggest inspiration came from small birds known as the Galapagos finches. When he disembarked on the Galapagos Islands, he noticed that the finches varied greatly from island to island in terms of their appearances, especially the beak form. Although islands were sometimes only a few miles apart, the differences were distinct, and they seemed to correspond to the differences in the environment. For example, on an island where droughts were more likely, plants produced fewer but larger seeds, so having a larger beak could be an advantage. Conversely, on islands with wetter climate, seeds were smaller, so a small narrow beak could do a better job of extracting them from various cracks. This resulted in more than a dozen species of finches unique to this remote archipelago.

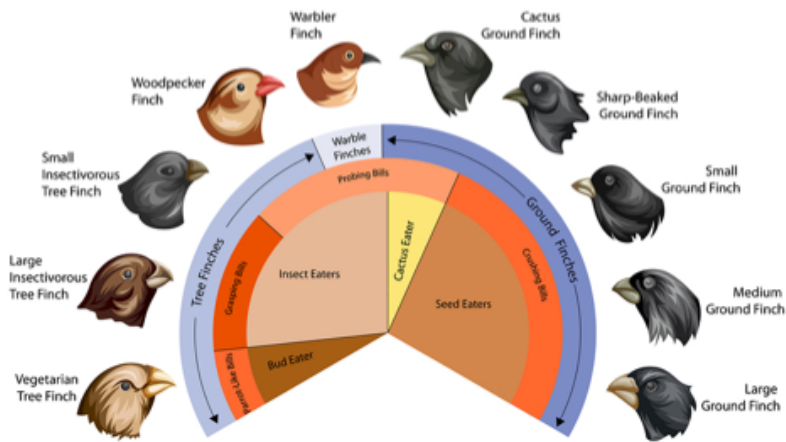


Image 10. A chart showing various adaptations in Darwin's finches

One thing to note is that in Darwinism, adaptation is driven by requirements of the environment. To rephrase this, adaptation is driven by the experiences that organisms have with the environment. If extracting seeds from tiny cracks between stones is a part of your experience sample, then the shape of your beak becomes important and your survival depends on it.

Evolution of personal beliefs

Can we extend the logic of Darwinian evolution to development of personal knowledge? Suppose beliefs are "organisms" that need to adapt to a certain environment in order to survive, while beliefs that do not fit too well quickly die out.

I can see pros and cons in this idea, but before criticizing, let's give it a try. I hereby present to you a "theory of Darwinian evolution of personal beliefs":

- Growing up, we develop an array of different personal beliefs (some from parents, some from media, some from education and other sources). This is *natural variation*.
- The environment we live in provides us with an experience sample. Of all the experiences we could possibly have, we actually have only a really small subset. Some of our beliefs are better fit to this experience sample. This is *differential fitness*.
- We test the beliefs against our experiences, and those that do not provide a good fit gradually fade. This is *survival of the fittest*.
- Beliefs that have survived form into complexes and produce new, related beliefs. This is *adaptation through natural selection*.

An example

Imagine that a friend told you that, according to an old belief, if you want to attract good luck at an exam, you should place a coin inside a shoe that you are wearing. You found it silly. On exam day, several of your classmates had coins in their shoes and they seemed to get satisfactory results, while you did not have a coin to protect you and your results were not satisfactory at all. You still thought that was a mere coincidence, but as the next exam day approached,

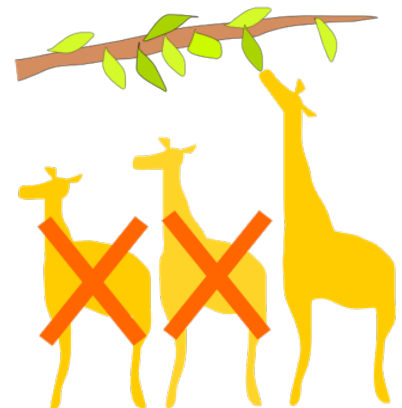


Image 11. Natural selection: those better fit to the environment have a higher chance of survival (credit: Toony, Wikimedia Commons)

Are superstitions a form of knowledge? (#Perspectives)

you thought there would be no harm in this tokenistic act so you placed a coin in your shoe and got good results. Since then, you never had an exam without a lucky coin in your shoe. That is how your experiences shaped a *superstition*.

In all probability, there were many various beliefs that potentially could become a superstition, but only one of them survived because it provided a good fit to your experience sample.

Conclusion

To sum up, it looks like we can draw an analogy between the development of personal beliefs and the development of species in Darwinian evolution. When we use the term “Darwinian evolution of personal knowledge”, we imply that the dependence between personal experience and personal knowledge is analogous to the dependence between the environment and natural selection of species.

KEY IDEA: We can draw an analogy between the development of personal knowledge and Darwinian evolution of species

If we accept this analogy, it opens up many interesting implications. These will be explored in the following lessons.

Critical thinking extension

At the start of this lesson, I posed the question: “How exactly do personal experiences influence the development of personal knowledge?”

I suggested the following answer: “They influence the development of personal knowledge in a process analogous to Darwinian evolution.”

Here are a couple of other questions on the relationship between personal experience and personal knowledge that I find interesting:

- 1) If we know what someone’s experiences have been, can we predict this individual’s beliefs?
- 2) Is it at all possible to break free from the prison of your experiences and have beliefs that transcend them?

If we accept the analogy with Darwinian evolution, how do you think we should answer these questions? Try to formulate an answer before you read on.

Here is my suggestion:

- 1) Yes, we can, but only to a very small extent. If I give a biologist a description of some environment, will they be able to describe what kind of creatures should have evolved in this environment? It should be possible to some extent. For example, if the environment is a desert, we know that the organisms should have developed some mechanism to survive extreme heat. That being said, exact predictions are probably not a possibility. Imagine I gave an alien biologist (who has never visited Earth) a thorough description of a desert. Will the alien biologist be able to draw a camel?
- 2) No. There is nothing in the process of Darwinian evolution that transcends the requirements of the environment. A Galapagos finch cannot just sit back, reflect and decide that it wants its future generations to develop narrower beaks. The only way to a narrow beak lies through wetter climate and smaller seeds. This conclusion is disappointing, isn’t it? But if you agree that development of personal knowledge is a Darwinian process, you will also have to accept this conclusion.

Can personal beliefs be fully predicted from personal experiences? (#Perspectives)

If you are interested...

Darwin's theory is so influential that it has been applied to a wide range of phenomena. The concept of **Universal Darwinism**, which emerged in the course of time as a summary of these applications, suggests that evolution through natural selection can occur in the world of non-living as well as living things. Some even tried to apply Darwinism to the development of the Universe.

If you are interested to know more, check out Lee Smolin's book *The Life of the Cosmos* (1997). In this book, he hypothesizes about cosmological natural selection and suggests that black holes, when they collapse, give birth to new universes on the "other side". In these new universes, the starting physical parameters are reshuffled a little (in a process analogous to "reshuffling" parents' genes in an offspring). As these universes develop, some of them are more successful than others. Obviously, these new universes also contain black holes that collapse and give birth to new universes, and so on.

Take-away messages

Lesson 3. There are some important similarities between the process of Darwinian evolution of species and the development of personal knowledge. They both seem to have the key features – natural variation, differential fitness, survival of the fittest and adaptation through natural selection. We may conclude from these similarities that the processes are, in fact, analogous. In other words, the development of personal knowledge is a Darwinian process. Once we accept this idea, it has some interesting consequences that are worth exploring. They will be further explored in the following lessons.

Lesson 4 - Analogy analysis

Learning outcomes

- [Knowledge and comprehension] What is the process of analogy analysis? What is false analogy?
- [Understanding and application] Can the analogy between development of personal beliefs and Darwinian evolution of species be considered false analogy?
- [Thinking in the abstract] What would be the Darwinian analogy for bias in personal knowledge?

Recap and plan

In the previous lessons I suggested an analogy between development of personal beliefs and Darwinian evolution of species.

Using analogy is in itself an important thinking tool; it is very popular but very tricky. **Analogical reasoning** is a valuable skill because it is so widely used in the production of knowledge in almost every area. Therefore, it is worthwhile to take a step back and formulate some general rules of analogical reasoning. In this lesson, using the analogy between development of personal beliefs and evolution of species as an example, we will look at the process of analogical reasoning and analyze the dangers of “false analogy”.

These are all transferrable metacognitive skills and concepts that you can use elsewhere.

False analogy

Analogies are a tricky thing because there exists a danger to fall for the so-called **false analogy** – it is a logical fallacy where the analogy is based on inessential characteristics (while the essential ones are different). Therefore, false analogies are misleading and result in flawed conclusions.

An example of a false analogy: *When a doctor is planning a surgery for a difficult case, it is okay to consult medical books. Therefore, medical students should also be allowed to use textbooks when they are writing exams.*

Obviously, there exists some similarity between surgeons conducting a surgery and medical students taking an exam. Both are stressed. Both are short for time. Both tasks are important. However, there are some essential aspects that are different: the goal of a surgery is to save a life, while the goal of a medical exam is to test your knowledge in order to later allow you to save a life. It is fine to consult books if you are lacking knowledge while conducting a surgery, but it is not fine for surgeons to rely on books by default.

False analogy ignores this crucial difference.

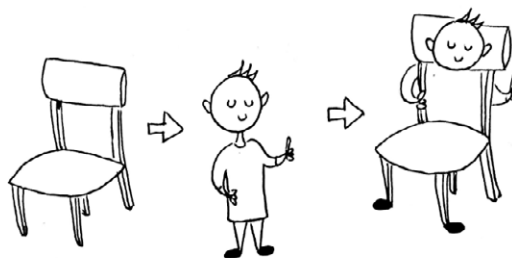


Image 12. False analogy: chairs have legs, I have legs, therefore I'm a chair

Key concepts

Analogical reasoning, analogy, false analogy

Other concepts used

Logical fallacy, essential characteristics and superficial characteristics

Themes and areas of knowledge

Theme: Knowledge and the knower

Does analogical reasoning provide sufficient justification for accepting beliefs as true?
(#Methods and tools)

Analogy between development of personal knowledge and evolution of species: false analogy?

Is the analogy I have drawn in the previous lesson a false analogy?

Is development of personal beliefs analogous to evolution of species?
(#Perspectives)

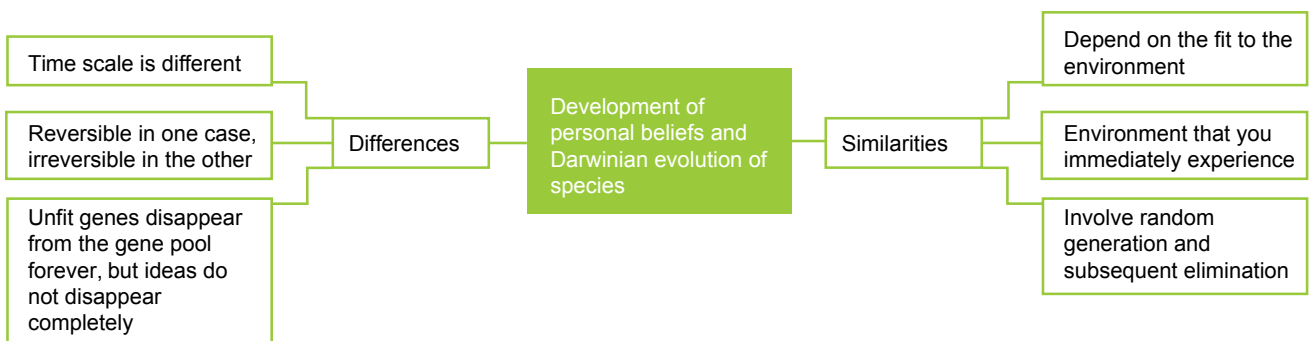
On the one hand, Darwinian evolution of organisms and development of personal beliefs do have similarities:

- 1) *Both depend on the fit to environment.* A Galapagos finch with a large beak will not survive in an environment where most food is hidden in narrow cracks. Similarly, the superstition about the lucky coin in your shoe is likely to fade if you get bad grades even when the coin is there.
- 2) *Both depend on the environment that you immediately experience.* It is only important to the bird what the cracks in the stones look like. Other aspects of the environment (for example, sea water temperature or the height of trees in the forest) are not essential because they do not influence the bird's everyday experiences. Similarly, when you develop a stereotype about inhabitants of a remote island, you base this stereotype on the experiences you have had: for example, the one article that you read or the one piece of gossip that you heard. It's hard to imagine how you can base your stereotypes on experiences that you might have had (but have not had).
- 3) *Both involve some random generation and subsequent elimination.* Mother Nature randomly generates offspring genotypes from the parents' genes. These genotypes then get tested against the environment and eliminated. Similarly, we have a whole range of beliefs, ideas, misconceptions, perceptions and transient thoughts. Not all of them stick around for a long time. Those that do not get a favorable response from the environment are doomed for oblivion.

On the other hand, there are essential differences.

- 1) The time scale is different. Darwinian evolution of organisms happens over the span of millions of years. Development of personal beliefs is a matter of one lifetime.
- 2) In natural selection, unfit genes disappear from the gene pool. In development of personal beliefs, the beliefs themselves may be suppressed or forgotten, but (thankfully!) they don't completely disappear.
- 3) Natural selection of genes is not quite the same as "natural selection of ideas". When genes die out (due to a poor fit to the environment), they cannot really be reborn. We lost our tails long ago, and it is hard to imagine that one of us could have a child with a tail. With ideas, it is different. I can retrieve a long-forgotten idea no matter how much time has passed since I abandoned it. Ideas are never truly and irreversibly dead.

What would your judgment be? Does the analogy stand? Can we dismiss the differences as inessential?



I personally think we can, making the analogy not false; however, I can also see why many people will not agree with me. No matter what you decide, the lesson here is that before deciding whether an analogy is true or false, you first need to decide which aspects of the two things you are comparing are essential and which aspects are more superficial. An analogy is only true if it is based on similarity in *essential* aspects.

How can we decide if the differences between two phenomena are essential or superficial? (#Methods and tools)

KEY IDEA: An analogy is only true if it is based on similarity in essential aspects

This raises a question: which aspects of an object or a phenomenon are considered essential? While there is no simple answer to this question, think about it in the following way:

When you take away an essential aspect, A is not A anymore. When you take away an inessential aspect, A may take a different form, but it still remains A.

For example:

Being a mammal is an essential aspect of a cat, but being furry is not. A cat that is not furry is still a cat (in fact, some breeds of cats look more like snakes, if you ask me). However, a cat that is not a mammal is not a cat. It is something else.

Thinking tool: analogy analysis

Let us formalize some rules of analogical reasoning in a concise form, so that you can use these tools elsewhere in thinking about knowledge.

Analogical reasoning is when you:

- 1) Observe that A and B are similar in essential aspects
- 2) Claim that A and B are analogous
- 3) Hence, infer that A and B must be also similar in all other aspects

For example:

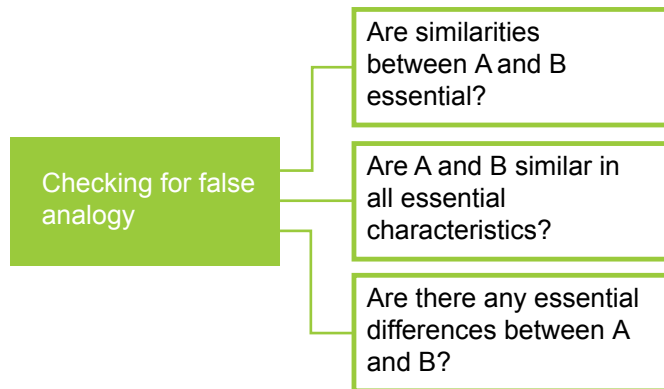
- 1) You observe that chimpanzees and humans are biologically similar in many ways
- 2) You claim that chimpanzees and humans are analogous in terms of how they respond to treatment
- 3) Hence, you infer that drugs that appear to be effective to cure disease in chimpanzees should also be effective to cure disease in humans



Image 13. Analogical reasoning is based on seeing a similarity between two things

One needs to be cautious in using analogical reasoning because analogy often turns out to be false. To carry out analogy analysis, you should ask yourself the following questions:

- 1) Are similarities between A and B essential or superficial? In the context of the example above, the fact that both species have two hands, two legs and one head is probably superficial while the fact that the genotype is 96 percent identical could be essential.
- 2) Are A and B similar in *all* essential characteristics or only some of them? Analogy is only reliable if *all* essential characteristics are similar.
- 3) Are there any essential differences between A and B? For example, if the 4 percent of the DNA sequence that is different between humans and chimpanzees codes for the immune system, it could actually be very essential. It could cause the reaction to drugs in the two species to be entirely different.



Analogical reasoning is powerful, but, to use it correctly, you need to make sure that you are not falling victim to false analogy (a logical fallacy). To do that, carry out analogy analysis! Use thinking tools to think better.

Critical thinking extension

The focus in this unit is the concept of bias. If development of personal beliefs and Darwinian evolution of species are indeed analogous, what counts as bias in these two processes?

What is the role of bias in the evolution of personal knowledge? (#Scope)

At the start of the unit, we defined bias as a systematic deviation (from some standard or truth). It seems to be easy to apply this concept to personal beliefs. A personal belief is biased when it deviates systematically from some “truth”. For example, my belief about the islanders of Tristan Da Cunha is biased if it does not correspond to the real state of things (we can go and check and see if the belief was biased or not).

However, what about the process of evolution? What counts as bias there?

If you are interested...

In *Monty Python and the Holy Grail*, the 1975 British comedy, there is an episode demonstrating the dangers of false analogy. Old, but still highly relevant today! You can watch the relevant episode in the video entitled “Monty Python deductive reasoning” on the YouTube channel *RegieNetCom110*.



Take-away messages

Lesson 4. Analogical reasoning is a thinking tool that is widely used in the production of knowledge in various areas. One observes that A and B are similar in some essential aspects and concludes that therefore A and B must be similar in other aspects, as well. However, when using analogical reasoning you should be cautious about false analogy. To ensure that the analogy is not false, one needs to decide if the characteristics that are similar in A and B are essential or merely superficial. The analogy between development of personal beliefs and Darwinian evolution of species seems to be based on some essential similarities, although there are also some differences.

Lesson 5 - Cultural experience

Learning outcomes

- [Knowledge and comprehension] How do people from different cultures differ in terms of sense perception, thinking and decision-making?
- [Understanding and application] What is the evidence supporting the claim that culturally specific experiences may influence the way we process information?
- [Thinking in the abstract] To what extent can we claim that culturally specific experiences shape culturally specific knowledge?

Key concepts

Culturally specific experiences, enculturation

Other concepts used

Sense perception, thinking and decision-making, trolley problem

Themes and areas of knowledge

Theme: Knowledge and the knower

Recap and plan

In the previous lessons we looked at how bias may be created in personal knowledge. I used an analogy with Darwin's evolution of species to claim that inevitable limitations of our personal experiences impose limitations on our personal beliefs.

If my claim that personal experiences shape personal beliefs is correct, then **culturally specific experiences** must also shape culturally specific knowledge. This is the claim that we are going to investigate in this lesson.

To what extent does culture influence personal knowledge? (#Perspectives)

KEY IDEA: Culturally specific experiences shape culturally specific knowledge

I am going to give you several examples of empirical evidence that support this claim. These examples will show how cultural experiences may influence:

- 1) Simple acts of sense perception
- 2) More complex cognitive phenomena such as thinking and decision making

Cultural experiences influence simple acts of sense perception

In the late 1950s, anthropologist Colin Turnbull spent time among the Bambuti Pygmies in the Ituri Forest in Congo, observing their behavior. He had a local 22-year-old guide, Kenge. In a fascinating series of stories, he describes how Kenge, who grew up in a thick forest and was never exposed to vast distances, travelled with Turnbull and saw prairies for the first time. They saw a herd of buffalo grazing on the plain a couple of miles away. Kenge turned to Turnbull and asked him what kind of insects they were. He lacked the mental machinery necessary to understand that large objects at a distance appear small. He just saw them as small objects. When Turnbull tried to explain this to his guide, Kenge, of course, didn't believe him so Turnbull drove to the buffalo herd. As the insects started rapidly increasing in size, Kenge asked what kind of witchcraft was involved (Turnbull, 1961).



Image 14. Grazing buffalo: they seem small if you look at them from a distance

In the famous Muller-Lyer illusion, you are required to say which of the two lines appears longer – the top one with the feathers turned inwards or the bottom one with the feathers turned outwards.

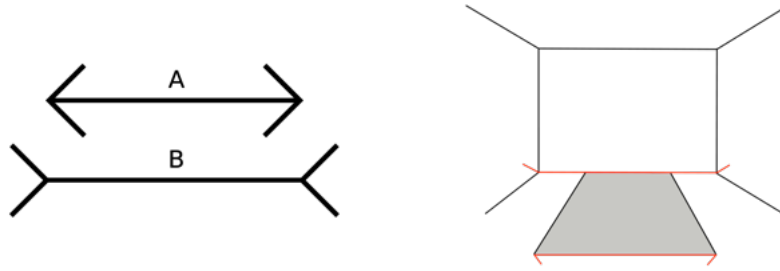


Image 15. Muller-Lyer illusion and our experience staying indoors

Most people say that the bottom line appears longer (although in reality they are the same length). It seems to be a universal phenomenon, an illusion built in the circuitry of our brains.

Is there anything culture-free in personal knowledge? (#Scope)

It is not for indigenous peoples of the Torres Strait Islands (situated between Australia and Papua New Guinea). When the anthropologist W.H.R. Rivers offered this test to the locals, he found that they were not susceptible to the illusion (Deregowski, 1998). Later he found that the same was true about many other indigenous peoples not exposed to the advances of civilization, such as the Toda people of southern India and the San people of the Kalahari Desert. An explanation that he suggests is that people in these pre-modern societies do not stay indoors as much as we do, and even when they are indoors, they are not surrounded with as many rectangular objects. Think about it: our houses are rectangular, our rooms are rectangular, our furniture tends to be rectangular. In such surroundings, if the angles along the edge of an object are out, the object is farther away from us. If the angles are in, the object is closer to us. Our human brains are not hard-wired to be susceptible to the Muller-Lyer illusion... well, they are, but only if we live in a modern society and stay indoors often.

Cultural experiences influence thinking and decision-making

In one study, participants were given tests where each question consisted of three pictures (such as cow, chicken and grass) and the task was to select the odd one out. It was found that American students (grades 4-5) consistently group objects based on belonging to a certain category – for example, they said that cow and chicken go together because they are both animals and that grass is the odd one out. By contrast, Chinese students of the same age consistently grouped objects on the basis of contextual commonality – for example, they grouped cow and grass together (because cows eat grass) and named chicken as the odd one out (Chiu, 1972).



Image 16. Cow, chicken, grass – which one is the odd one out?

So... your culture determines how you think? It seems plausible. After all, we gradually absorb all the aspects of our culture as we are growing up (this process is called **enculturation**). It probably means that if you are exposed to several cultures when you

are growing up (a multicultural environment), your thinking will be more flexible. One is tempted to believe so. What if the reality is that the two (or more) cultures do not mix up and enrich each other, but instead reside in your mind as independent entities and you switch between them from time to time? That would be simultaneously awesome and spooky!

If you are bilingual, it turns out the language you are speaking at the moment influences the way you are thinking. Your answers to the same questions may depend on what language the question is asked in!

For example, research with university students in Hong Kong (who were fluent in both English and Cantonese and had considerable exposure to both cultures) showed that in various decision-making scenarios such as deciding which camera to buy or which restaurant to go to, participants were more likely to make compromise choices and avoid potential disappointment when speaking Cantonese. When instructions were presented to the same students in English, their decisions became much riskier and more extreme (Briley, Morris & Simonson, 2005).

Another researcher found that when bilingual individuals are presented with a moral dilemma, they tend to make emotion-driven decisions when the dilemma is presented in their native tongue and more logic-driven decisions when the dilemma is presented in the second language that they speak with more effort. For example, in one of the modifications of the “trolley problem”, there is a train going towards five people working on the tracks that is about to kill them. You see this from a bridge above the tracks. On the same bridge, there is a fat man. You know that if you push him down from the bridge, he will get killed but his body will slow down the train and prevent the death of five people. The question is, are you willing to push the fat man from the bridge in order to kill one but save five? Apparently, if the dilemma is given to you in your second language, you are more likely to say yes (Costa et al., 2014).

Should judgments of morality of an action depend on the context in which the action is taking place?
(#Ethics)

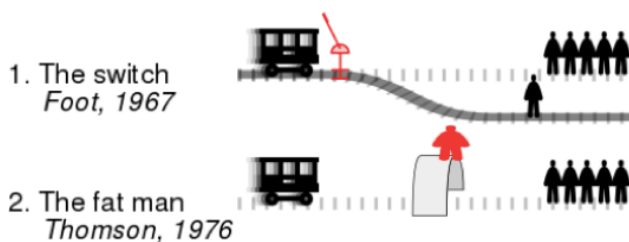


Image 17. The trolley problem and its modification with a fat man on the bridge (credit: Cmglee, Wikimedia Commons)

This research with bilingual individuals also seems to suggest that the two cultures do not integrate in our mind in one holistic entity. Instead, they seem to continue to co-exist as two independent and self-sufficient entities, and you activate either of them depending on the situation, such as when a particular language is being spoken.

Conclusion

Research studies reviewed in this lesson (as well as tons of research studies that are beyond the scope of this book) suggest that culturally specific experiences may influence the way we process information. Apparently, this happens on many levels, from the simplest acts of perception to rather complex acts of thinking and decision-making.

KEY IDEA: Culturally specific experiences may influence the way we process information

Critical thinking extension

After reviewing some empirical evidence, the conclusion we have arrived at is “Culturally specific experiences may influence the way we process information”.

However, the key argument I put forward at the start of the lesson is “Culturally specific experiences shape culturally specific knowledge”.

Do you feel the difference? As with everything in Theory of Knowledge, let us be reasonably skeptical about our statements. I invite you to contemplate the following questions and arrive at your own conclusions:

- 1) Is information processing the same as knowledge? If we process information differently, does it necessarily mean that personal knowledge we arrive at is also different?
- 2) Reiterating the conclusion, culturally specific experiences *may* influence the way we process information. They may – but do they always? Is it possible, for example, to grow up in a culture but consciously override the influence of this culture on some aspects of your thinking?
- 3) To what extent are these cultural differences essential? Can we claim that cultural differences in information processing (and personal knowledge?) are so large that people from different cultures will not understand each other on a deep level? Or are these differences negligible?

Once you understand that culture influences personal knowledge, can you override this influence with rational thinking?
(#Methods and tools)

If you are interested...

Taking one step further, there is also evidence that cultural experiences influence the structure of our brain! In other words, brains of people from different cultures are also somewhat different. You can read more about this here: Park, D.C., and Huang, C.-M. (2010). Culture wires the brain: A cognitive neuroscience perspective. *Perspectives on Psychological Science*, 5(4), 391-400.

Take-away messages

Lesson 5. Cultural experiences may shape the way we process information. This is evident on many levels, from simple acts of sense perception to complex acts of thinking and decision-making, including ethical reasoning. This influence exists because we gradually absorb various aspects of our culture as we are growing up, in a process known as enculturation. Research with bilingual individuals suggests that enculturation to several cultures at once creates separate, independent “modules” of information processing in our minds, something like several minds within the same person. A stronger claim based on such evidence would be to say that culturally specific experiences shape culturally specific knowledge.

Lesson 6 - Memes and Universal Darwinism

Learning outcomes

- [Knowledge and comprehension] What is a meme? How is a meme similar to a gene?
- [Understanding and application] To what extent does memetics apply to the development of personal knowledge?
- [Thinking in the abstract] Is free will merely an illusion created by memes that disguise themselves as the host's own ideas?

Recap and plan

We already used the logic of evolutionary theory to explain how our personal knowledge may depend on the personal experiences we have been exposed to. Personal knowledge may be an instance of adaptation to the requirements of the environment.

In this lesson, rather than just exercising a critical comparison between two phenomena, we will consider a formal theory that already exists. I will introduce the concept of memes and the field of study known as memetics. Memetics is one of the products of Universal Darwinism – the idea that principles of evolution apply universally and not only to natural selection of biological species. I must say that the idea of memes is not fully accepted in academic circles, but it is still worth considering for the sake of raising interesting questions about bias in personal knowledge.

I will explain the concept of memes and give you a gist of the main ideas of memetics. Just a heads up: by the end of this lesson, I will claim that you only exist as a host for the spread of a cultural virus, that you are merely a vessel devoid of free will, that your self is an illusion that the virus has created to make you more complacent to its influence. Well, you know, the mundane reality of life.

What is a meme?

Richard Dawkins is a very popular evolutionary biologist and a prolific writer. In his first bestseller, *The Selfish Gene* (1976), he introduced a gene-centered view of evolution.

The main message is as follows:

- 1) It is not survival of the organism that drives evolution, but survival of a separate gene.
- 2) Hence it does not matter to the gene if it is passed to further generations by its host organism or by some other organism, as long as it gets passed on.
- 3) This explains many instances of selfless behavior that can be observed in various species. Organisms will sacrifice themselves to increase chances of survival of other organisms, but only if the two organisms are genetically related.
- 4) Therefore, your survival only matters as long as you maximize the chances for your genes to replicate. If there's a better way for your genes to replicate (for example, you sacrificing your life for your brother who carries a similar genotype), your genotype will not think twice.

Key concepts

Meme (a unit of culture that bears a certain meaning), memetics, Universal Darwinism, analyzing implications

Other concepts used

The selfish meme, replication, variation, differential fitness

Themes and areas of knowledge

Theme: Knowledge and the knower

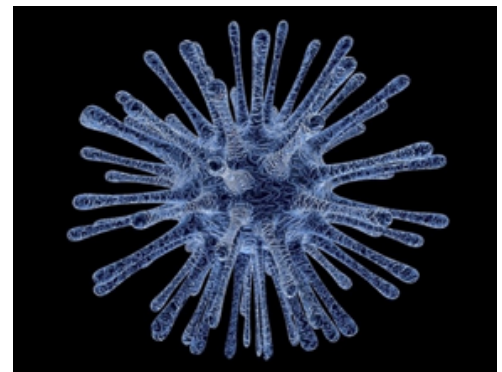


Image 18. Viruses spread from one host to another

If you ask me, the implications of this are a little scary. The way I see it, there exists a whole parliament of little voters inside my body (we have an estimated 20,000 genes) who are very interested in replicating themselves. Each one of them has relatives living in the bodies of a bunch of other people. They protect the interest of their relatives. In every particular situation, they take a call on what behavior would be best to achieve this purpose. If as a result of this behavior my personal life is at risk, they don't really care! As a separate organism with its dreams and desires, I am actually pretty inessential.



Image 19. Are genes selfish?

But there's more.

Is the concept of a meme a false analogy? Is it justified to speak about a "unit" of knowledge in culture? (#Perspectives)

In the last chapter of his 1976 book, as an extension of his ideas on biological evolution, Dawkins introduced the term **meme**. Just like a gene is a unit of heredity coding for a specific observable trait (eye color, height, lactose intolerance), a meme is a unit of culture that bears a certain meaning (a catchy tune, the idea of God, a ritual, a greeting sign). Just like collections of genes that code for some complex trait (for example, a collection of genes that determines if you will make a good soldier), memes can be combined in complexes – called **memeplexes**. Examples of memeplexes include religions, languages, works of art.



Image 20. Spreading of memes

What are the necessary and sufficient conditions for evolution of knowledge? (#Scope)

Dawkins is a proponent of **Universal Darwinism** – the view that evolution is not limited to the biological world, that evolution must occur in any other situation where the three conditions are met. For example, it may apply to the first self-replicating molecules. An important point to note here is that the three conditions are *necessary and sufficient*: without any of them, evolution will not occur, but if all three are present, evolution *must* occur.

Universal Darwinism

According to Dawkins, evolution will occur whenever three conditions are met:

- a) replication,
- b) variation,
- c) differential fitness.

KEY IDEA: Universal Darwinism:
replication + variation + differential fitness = evolution

Applying Universal Darwinism to memes

Just like a gene, a meme can replicate itself. “Vertical” replication is from generation to generation (for example, parents teaching their children that they should not trust strangers). “Horizontal” replication is within one generation from one person to another. An example of “horizontal” replication is a video that becomes viral on



Image 21. The DNA has an ability to self-replicate

YouTube and spreads across the world. The idea of a meme is itself a meme, and at the moment I am engaging actively in its horizontal replication.

Just like there exists a natural variation of genes in the gene pool, there exists a natural variation of memes in the meme pool. You can feel that very evidently these days when you log in to Netflix and start choosing a TV show to watch in the evening. There is so much stuff available that it is really difficult to choose sometimes.

Just like genes show differential fitness, some memes survive and replicate better than others. Do I need to explain that? J.K. Rowling's *Harry Potter* was very successful, but *Firefly*, a sci-fi show I really enjoyed, was cancelled after the first season.

Since all three conditions are met, evolution has to occur. The idea is that memes, just like genes, evolve through a process of natural selection.

Memes and bias in personal knowledge

Now how does it all link to bias in personal knowledge? For things to evolve, there needs to be variation. What we might perceive as “biases” in personal knowledge – stereotypes, misconceptions, strange beliefs and superstitions – may simply be instances of this variation. A meme, from the evolutionary viewpoint, is an instance of trial-and-error. Just like all sorts of weird creatures exist in the biological universe because natural variation produces them to see who survives, all sorts of weird biases in personal knowledge exist because natural variation of memes produces them – for some of them to survive and for some to die out.

Is bias a necessary condition for the development of knowledge?
(#Methods and tools)

KEY IDEA: Natural variation of memes is a necessary condition for their evolution. If personal beliefs are memes, then bias in personal knowledge is necessary because it enables natural variation.

So... which memes do you host? How actively are you passing them on? How likely do you think it is that these memes will survive?

And has the “meme meme” successfully replicated itself by jumping from my mind to yours?

Critical thinking extension

The selfish meme

As the “meme meme” took its roots, scholars have developed this idea resulting in the emergence of a whole new field of study – **memetics**. One of the famous founders of this movement is Susan Blackmore, with her bestselling book *The Meme Machine* (2000).

Just like the idea of a selfish gene implies that survival of the organism carrying genes is not as important as survival of the genes themselves (remember the 20,000-member parliament within your body?), the idea of evolving memes implies that every particular individual is nothing but a host. A vessel for a virus.

Let us take this idea and explore its *implications*.

Is morality a cultural meme?
(#Ethics)

It provides an evolutionary advantage for a meme if its host thinks that he or she has free will, if there is an illusion that the meme was the host's conscious choice. For example, you might have seen the show *The Voice*. Chances are, it exists in your country. Chances are, you have seen your local version of *The Voice* and perhaps the US version, but not any of the episodes from the rest of the 145 countries where the show has been adapted! When you log on YouTube and look for *The Voice* episodes, the search algorithm politely provides episodes from your own country. It is actually very likely for people to not even suspect that the show exists somewhere other than in their native country. *The Voice* meme maximizes its chances of survival (being seen and being passed on) if it pretends like it is unique for every specific cultural group.

Could it mean that we don't choose our beliefs, but rather our beliefs choose us? They choose us but they also manipulate us into thinking that it was us who chose them – this way they will survive longer.

If you are interested...

If you are interested in studying memetics more closely, a wonderful introduction is Susan Blackmore's book entitled *The Meme Machine* (2000). You can also visit the author's website: www.susanblackmore.uk/

Take-away messages

Lesson 6. Universal Darwinism is the idea that the process of evolution is not limited to natural selection of biological species, but must occur whenever three conditions are met: replication, variation and differential fitness. Memes are units of culture that bear a certain meaning. They get replicated both vertically and horizontally. According to the "selfish meme" idea, meme hosts are not as important as the memes themselves. Memetics provides a formal application of Darwinian evolution to the development of personal knowledge.

Lesson 7 - Heuristics

Learning outcomes

- a) [Knowledge and comprehension] What is a heuristic?
- b) [Understanding and application] What are the advantages and disadvantages of using heuristics?
- c) [Thinking in the abstract] What implications does the existence of heuristics have for our understanding of the way humans think?

Recap and plan

In the previous lessons we have been exploring bias in personal knowledge. We used the evolutionary theory to make sense of it. The take-away message is that personal knowledge may indeed be biased, but this bias comes from the natural variation of our personal experiences. In a process akin to natural selection, beliefs that provide an evolutionary advantage stand the test against our experiences and get reinforced (even if they are biased!).

In the next couple of lessons, let's look at how exactly this happens. In particular, we will look at what biases exist in our mental software and why they are still there despite understanding that they are biases.

System 1 and System 2 thinking

A lot of what we know today about biases in personal knowledge comes from psychology. Amos Tversky and Daniel Kahneman were the two scientists who framed this research as a systematic field of study and developed a coherent theory of **cognitive biases** (Tversky & Kahneman, 1974). The theory suggests that there are two “systems of thinking” that humans use when they process information and make decisions – System 1 and System 2 (Kahneman, 2011).

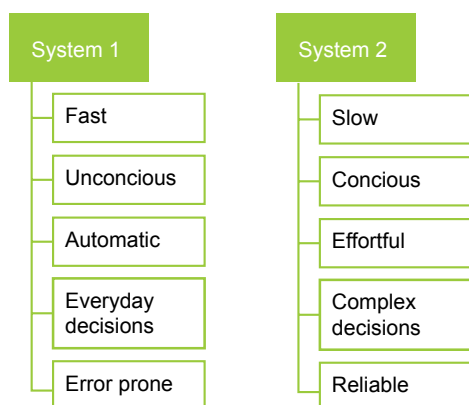


Image 22. System 1 and System 2: the intuitive and the logical systems of thinking

When you tell yourself that people are probably likely to overestimate dangers of travelling by air or when you compare death statistics from airplane crashes against car accidents and learn that planes are, statistically, much safer than cars to travel, that is your System 2 overriding the initial reactions of System 1.

Key concepts

Heuristics, cognitive biases, System 1 and System 2 thinking, normative and descriptive models of thinking

Other concepts used

Anchoring bias

Themes and areas of knowledge

Theme: Knowledge and the knower

How can personal knowledge be biased despite our awareness that it is biased? (#Scope)

System 1 thinking developed earlier in the process of evolution, and humans are not the only species that have it. System 1 is quick, automatic, intuitive and based on past experiences. When you see news about an airplane crash that claimed lives of people and you are afraid to fly because it seems to you that airplanes are a dangerous kind of transport, that is your System 1 speaking. It uses vivid perceptual images to make sweeping generalizations about things it does not completely understand. It is the cause of irrational behavior.

How reliable is knowledge that is a product of intuitive thinking? (#Methods and tools)

System 2 thinking is deliberate, logical, rational and analytical. It is a consequence of our

According to Kahneman (2011), System 1 and System 2 act sequentially. First, we react with our quick intuitive brain and then – if necessary – we override that reaction with our logical, rational, “educated” brain. This makes sense.

- First, if we were using System 2 constantly rather than occasionally, our life would be a nightmare. For every simple decision, we would be spending loads of time and energy and cognitive effort.
- Second, in most cases, System 1 works just fine. The logic that System 1 uses is “Look, I did this before and it worked, so I can do it again” and, yes, it will probably work again.
- Finally, since we evolved from more primitive animals, it makes sense that we have all the mental machinery they have, plus something on top of that. Evolution did not re-wire our brains completely; it wrote patches and created additional modules.



Image 23. First we think fast, then we think slow (credit: P.O. Arnäs, Flickr)

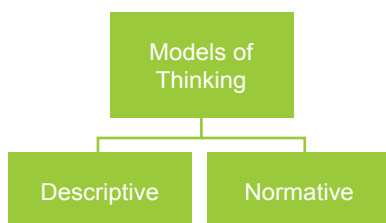
KEY IDEA: System 1 and System 2 act sequentially: first we use intuition, then we override it with rational analysis

Heuristics and cognitive biases

What knowledge is more valuable: descriptive knowledge of how people think or prescriptive knowledge of how people should think? (#Perspectives)

Models that describe System 2 thinking are called “**normative models**”. They are “normative” because they tell us what is correct and incorrect, which decision or conclusion is accurate and which is not. Examples of normative models include logic, utility theory, and probability theory. For example, probability theory may be used to arrive at the “normative” answer to the question “How dangerous is it to travel by air?”

Models of thinking that show the workings of System 1 are called “**descriptive models**”. They are named this way because they describe thinking as it is, not as it should be. Descriptive models are comprised of so-called heuristics and cognitive biases. **Heuristics** are “cognitive shortcuts”, simplified thinking strategies that we use under lack of time, incomplete information or similar restraints. These are utilized to save time and mental energy.



Since heuristics are based on past experience, much of the time they work fine (they are good enough). They worked in the past, so it is likely – to a certain extent – that they will work again. If they do work, there is no issue, but if they do not work, they result in cognitive biases. Cognitive biases are systematic deviations of thinking from what is dictated by normative models.

KEY IDEA: Since heuristics are based on past experiences, much of the time they work fine. But when they don't, they may result in cognitive biases.



Image 24. Cognitive bias: afraid to fly, although there is no reason

An important discovery in psychology is that cognitive biases are predictable. People make predictable mistakes in predictable situations, which is great news (for science, maybe not so much for people!).

For the sake of illustration, I will give you just one example from a pool of hundreds of cognitive biases that have been discovered – anchoring bias.

Anchoring bias

Anchoring bias occurs when you make a decision based on an initial piece of information (an anchor) provided to you, even if the anchor is not very relevant.

For example, suppose you are buying a used laptop. You ask what the price is and the seller says X. This X is the anchor. In the subsequent conversation if you settle on a price lower than X, it will seem like a good bargain, and if it is substantially lower than X, the seller will appear to be making sacrifices. This will happen, to some extent, even if X is actually higher than the market price. So it all depends on where X – the anchor – is initially placed.



Image 25. Anchoring bias: we use an “anchor” as a starting point in thinking about numbers

Strack and Mussweiler (1997) asked two groups of students whether Mahatma Gandhi died before or after age 9 (group 1), or before or after age 140 (group 2). Both of the anchors were quite ridiculous, and students in group 1 said “after”, while students in group 2 answered “before”. However, when these same students were asked to say at what age they thought Mahatma Gandhi died, the average guess differed significantly in the two groups (age 50 in the first group versus age 67 in the second group).

How can we ever know if our personal knowledge of something is biased? (#Methods and tools)

In a more dramatic example, Englich, Mussweiler and Strack (2006) used practicing judges as participants. They gave them a hypothetical scenario, and the judges had to answer a series of standard questions and arrive at a decision (a sentence). Part of the scenario was the severity of the sentence demanded by the prosecutor. Judges were told that for the sake of the study, this parameter would be determined randomly. Judges were asked to throw dice and take the resulting number as what the prosecutor demands.

Results of the study showed a correlation between the final sentence awarded by the judge and the number on the dice: the larger the number, the more severe the sentence. This is of course concerning, because severity of the punishment demanded by the prosecutor (the anchor) has nothing to do with how guilty the alleged criminal is. Moreover, judges in this study were aware that the prosecutor demands were determined at random – they threw the dice themselves!



Image 26. Can dice determine the severity of a court's decision?

Critical thinking extension

What implications does the existence of heuristics have for our understanding of the way humans think?

Once again, remember that it is important in TOK to identify implications of arguments. An implication is a logical consequence. Suppose you have formulated argument X. Implications of X are all the things that must be true if X is true. Practice this skill! Look at the three arguments below and formulate the implications of these arguments (I will give you some hints which you can use or ignore):

Argument	Implications	Hints
Heuristics result from experience.	?	This fits nicely into the formula that we discussed previously: personal knowledge is based on personal experience. Heuristics will only survive if they already worked sufficiently well in the past.
Heuristics have an adaptive function.	?	We have them because it is beneficial in some way. For example, using an anchor to adjust your thinking is simple yet <i>usually</i> good enough. It helps us make acceptable decisions quickly.
Heuristics are predictable.	?	This allows us to study heuristics scientifically. This also creates a curious situation: our minds are riddled with these glitches but we are aware of them. Although we are aware of them, we cannot simply choose not to use them.

If you are interested...

Two great books to learn about heuristics and cognitive biases are: *Thinking Fast and Slow* by Daniel Kahneman (2011) and *Predictably Irrational* by Dan Ariely (2008).

A list of cognitive biases to be amazed and impressed by can be found on the Wikipedia page "*List of cognitive biases*".

Take-away messages

Lesson 7. System 1 and System 2 thinking act sequentially. The first (automatic, intuitive) decisions come from System 1, which is based on past experiences and includes a range of simplified thinking strategies called heuristics. Heuristics may or may not lead to cognitive biases. System 2 thinking may override these automatic reactions using rational, precise analysis. However, System 2 cannot be used all the time because it requires a lot of mental effort. Models of thinking that explain how thinking should work (System 2) are called normative models. Models focused on how thinking actually works (System 1) are called descriptive models of thinking. There are multiple examples of documented heuristics – one of them is anchoring bias. Heuristics have an adaptive function. Heuristics are predictable.

Lesson 8 - Implicit bias and bias self-awareness

Learning outcomes

- [Knowledge and comprehension] What is implicit bias? What is bias self-awareness?
- [Understanding and application] How are implicit biases different from explicit attitudes?
- [Thinking in the abstract] To what extent is it possible to become aware of your own implicit biases?

Recap and plan

In the previous lessons we considered some examples of biases in thinking and decision-making. I hope I have convinced you that:

- People use heuristics (cognitive shortcuts) in their thinking and decision-making
- Because of this, people are susceptible to lots of cognitive biases

If we accept all that, we should also probably accept that bias in personal knowledge is inevitable.

Many of these biases are implicit. People might be confident that they don't have them when in fact they do (remember memes? this might be one of those dirty tricks memes are using to ensure their survival!).

Is it possible to have knowledge of our implicit bias? Can I at least know where I am biased, or am I doomed to be oblivious about it?

I will try to outline possible answers to these uncomfortable questions.

Implicit biases

Implicit biases are a special type of bias that stay below the level of conscious awareness. This means that implicit biases affect our thinking and behavior without us realizing it. In fact, on the level of conscious awareness, we may be certain that we are not biased when in fact we are. This makes implicit biases very powerful in terms of affecting our lives.

KEY IDEA: Implicit biases affect our thinking and decision-making, but we don't realize it

I will illustrate implicit biases with the example of *implicit prejudice*. In human sciences a popular way to explore implicit prejudice experimentally is through the so-called "shooter bias paradigm". In this procedure you are playing a video game where figures appear in random places on the screen at random times. Some of these avatars are those of majority groups and some are those of minorities; some figures are holding a gun while some figures are holding harmless objects. Your task is to quickly push a button to "shoot" those avatars that are holding a gun. This reminds me of a scene from *Men in Black* (1997) where Will Smith, as part of his pre-employment exam, had to shoot aliens in a simulation.

Key concepts

Implicit biases and explicit attitudes, bias self-awareness

Other concepts used

Shooter bias paradigm, self-report questionnaire, implicit prejudice

Themes and areas of knowledge

Theme: Knowledge and the knower
AOK: Human Sciences

Is it possible to have knowledge of our own implicit bias? (#Scope)

Research studies using the shooter bias paradigm have demonstrated that it is common for people to show implicit prejudice in these simulations. For example, one study with American participants showed that black avatars holding harmless objects had a higher chance of being shot than white avatars holding harmless objects. Decisions to shoot were also made faster for black avatars than for white avatars (Correll et al., 2007). It looked like the brain quickly and automatically associated “black” with “dangerous”. In another study, the same findings were obtained for Caucasian avatars wearing Muslim headwear versus Caucasian avatars wearing no headwear. The brain appeared to associate “Muslim” with “dangerous” (Unkelbach, Forgas & Denson, 2008). Interestingly, these are all instances of implicit prejudice – they only show in a computer game simulation where participants have to make quick decisions. On the level of **explicit attitudes**, if you give these same participants self-report questionnaires, they indicate sincerely that they believe they are not prejudiced. How can they think they are not prejudiced when in fact they are?

To what extent are we responsible for putting effort into overriding our implicit biases? (#Ethics)

This could be explained by System 1 and System 2 acting sequentially. System 1 is irrational and automatic. It operates on vivid images that it gets from mass media and everyday experiences. If my only exposure to Muslims is (sadly) the very vivid pictures I have seen several times on TV in connection with suicide bombings, then my System 1 is very likely to operate on those images when making quick, automatic decisions. I am an educated person living in an educated society, so my System 2 will intervene and override this automatic reaction. However, the System 1 reaction remains the default one and requires some extra effort to override. In situations where there is no time to think (like the shooter bias paradigm), System 1 is used.



Image 27. Open mind switch: if only it was so easy

Some questions that emerge are: What do I need to do in order to become aware of my automatic System 1 reactions? What do I need to do in order to change them? If changing them is too difficult or impossible, what do I need to do in order to ensure that my System 2 intervenes more quickly, more reliably, more consistently? In other words, how do I eradicate this prejudice from the depths of my mind?

Bias self-awareness

Let’s switch from a serious example to a less serious example.



Image 28. Greek “Know Thyself” mosaic found at an excavation in Italy

Suppose you are implicitly prejudiced against *Harry Potter* in favor of *Lord of the Rings*. In other words, according to your implicit attitude, *Lord of the Rings* is much cooler. However, since it is an implicit attitude, you have no idea about it. On the contrary, explicitly you believe that *Harry Potter* is the coolest thing ever, and on many occasions you have agreed with your friends that *Lord of the Rings* does not live up to this standard.

Can you ever *know* that you are implicitly prejudiced against *Harry Potter*?

If this implicit attitude really exists, it is likely to affect your behavior in certain ways, but these effects will be masked by explicit attitudes. For example, on one particularly lonely evening when you had to watch something to kill

time, you decided to re-watch *Lord of the Rings*. You explained the choice to yourself by saying, “I need to refresh my memory to have better arguments and examples for the next time I need to convince people that *Harry Potter* is clearly better.”

In this situation you have an explicit belief (“I prefer *Harry Potter*”) and an explicit behavior (“I am watching *Lord on the Rings* on a lonely evening”). The trick is to notice your own explicit behavior and hypothesize about the existence of an implicit attitude that might explain it (“Why am I watching *Lord of the Rings*? Could it be that I actually prefer it to *Harry Potter*?”). Implicit biases are implicit, so you cannot just see them directly! To confirm your suspicions, you will need to observe your behavior a little longer (“Let’s wait until the next weekend and see what movie I will be in the mood for”) and even experiment with your own behavior (“Let me go to a *Lord of the Rings* fan’s party and see if I feel comfortable there”). From all of these clues, it might be possible to infer that you have an implicit attitude.

I find it genuinely amazing that to uncover my own implicit attitudes, I have to experiment with myself like it is not “me” but some other person that I barely know. However, it makes sense: if implicit biases exist, then there is a part of “me” that is inaccessible to my conscious self. In a way, there is a whole other person inside my mind who wants to influence my behavior but remains hidden.

We need a term for this ability of a person to be aware of their own implicit biases. We will call this ability **bias self-awareness**.

What is the most effective way to increase bias self-awareness? (#Methods and tools)

KEY IDEA: You can’t become directly aware of your implicit biases. You need to infer their existence from observing your own behavior.



Image 29. Self-awareness

Critical thinking extension

To what extent is it easier to recognize bias in others than in oneself?
(#Perspectives)

You will probably agree that bias self-awareness is a desirable trait, but to what extent can you train it?

Essentially, training is exposing yourself to new experiences. We know from the previous lessons that personal knowledge is dependent on personal experience. Bias self-awareness is part of personal knowledge. In order to train it, you should systematically expose yourself to special experiences where this ability will be engaged and developed.

The question is, what exactly are these experiences? In what situations do you think bias self-awareness becomes particularly necessary? If you can identify such experiences, you can change your daily routine in a way that will allow you to develop unprecedented bias self-awareness!

In the *Lord of the Rings* example above, what steps would you take to train your bias self-awareness?

Being self-aware about your biases may be a huge step towards being bias-free (although arguably this goal cannot be fully attainable).

If you are interested...

If you would like to test yourself on potential implicit biases, try taking several IATs (implicit-association tests) on Harvard's *Project Implicit* website.

Please make sure to read the instructions carefully.



Take-away messages

Lesson 8. Implicit biases are not accessible on the level of conscious awareness, but they affect our thinking and behavior. Explicitly, a person may be certain that they are not biased when in fact they are. This makes implicit biases very difficult to detect, both in other people and in yourself. The ability to be aware of your own implicit biases is called bias self-awareness. At least theoretically, bias self-awareness may be trained. This requires systematic effort and exposing oneself to new experiences.

Lesson 9 - Bias reduction

Learning outcomes

- [Knowledge and comprehension] What are the strategies that could be used to reduce bias?
- [Understanding and application] What are arguments for and against bias reduction in the acquisition of knowledge?
- [Thinking in the abstract] What is the role of reflexivity in bias reduction?

Recap and plan

We have discussed how, when it comes to personal knowledge, people have lots of biases built into their mental software. We have also seen that many of these biases are implicit; this means that they affect our decisions even though we are confident that they don't.

As discussed in the previous lesson, to some extent it is possible, at least theoretically, to increase your bias self-awareness. To do so, you should actively explore your own thinking and decision-making as if you were studying another individual, conduct experiments and test hypotheses about your own thinking.

Hopefully, being self-aware about your implicit bias may help you to reduce it to some extent and perhaps even eliminate it? In this lesson we are going to investigate the extent to which such **bias reduction** is possible.

What if we were bias-free?

What would it be like to be bias-free? To have the superpower of seeing the world and every single detail in it with unbiased, neutral, objective eyes? Would you like to have this superpower? Would you be happy if you were the only person on Earth to have this ability? Would you be happy if all humans suddenly became bias-free?

That is just a **“what-if” thought experiment**. Such thought experiments are a powerful thinking tool because through hypothetical scenarios you can explore dimensions of an idea that you cannot explore otherwise.

Think about these questions for a while. I would like you to formulate a response in your mind. Whatever the mental path you took, the destination that you reach is probably one of two things: either “It would be nice if we could be bias-free” or “It would be a disaster, better continue being biased”. Which of the two destinations have you reached?



Image 30. What-if thought experiment: what if snails had legs? (credit: Fishhead, Sketchport)

Key concepts

“What-if” thought experiment, bias reduction, counter-stereotypical information, reflexive control, reflexivity

Other concepts used

Bias-free individual, debiasing

Themes and areas of knowledge

Theme: Knowledge and the knower
AOK: Human Sciences

Is it possible for implicit bias to be eliminated through self-awareness? (#Scope)

“What-if” thought experiments

In a “what-if” thought experiment, you imagine that one aspect of this world is different from what it is, and then you logically derive what other aspects would be different. This may be an improbable situation. Examples include questions such as: What if you were immortal? What if there was no moon? What if the north pole and the south pole were swapped?

Can credible knowledge be acquired through a thought experiment? (#Methods and tools)

What-if thought experiments are a powerful thinking tool. They allow you to explore scenarios that are not accessible to everyday perceptual experiences. Although this may all seem highly hypothetical, the conclusions may be really eye-opening sometimes.

A great and funny resource that is tastefully written on such thought experiments is Randall Munroe’s book *What If?: Serious Scientific Answers to Absurd Hypothetical Questions* (2014).

Although I am hesitating, I think the destination I am reaching, sadly, is the second one – “better continue being biased”. Here are just a couple of arguments, without any intention to talk you into taking my side:

- 1) To be bias-free means to lose identity. Our biased opinions often rest upon groups that we belong to. For example, a biased historian may tweak their interpretations of events of the past because they (implicitly) want their nation to look good. Such bias is not a good thing. On the flipside, identifying yourself with a group and being impartial about it do not go together well. Although we are blaming the historian for their biased approach, they are being biased out of a sense of identity. Without identity, the lives we are living may be quite meaningless.
- 2) To be bias-free means to lose passion. When situations are uncertain and information is incomplete, we (biased individuals) form opinions. Since these opinions are ours, we dearly protect them. Trying to support an opinion may be a powerful driver of research and inquiry. On the contrary, to be bias-free probably means to have no opinions. Bias-free individuals will not try to prove anything to each other, so they will not be motivated to do research. Absence of bias may slow down progress in the acquisition of knowledge.
- 3) To be bias-free means to lose confidence. When a person is biased, they are overconfident in a belief that is not true. Without a doubt, in many situations being overconfident is a bad thing. Even so, being confident allows us to act. We live in a world that is full of ambiguity and uncertainty. If we do not jump to (biased) conclusions, we may find ourselves in a knowledge vacuum, without beliefs or values to stick to. We would doubt too much and do very little.



Is it better for a knower to be biased or bias-free? (#Perspectives)

For these reasons, my vote goes to “continue being biased”. Biased opinions are very valuable, I think. Even if I was given a chance to eliminate bias, I would not use it.

However, I still think that controlling – not eliminating – our bias to some extent would be nice.

KEY IDEA: Completely eliminating bias from personal knowledge may be undesirable

To what extent can we control our implicit biases?

Research existing in this area suggests that it is possible to control our implicit biases, to some extent. Some strategies attempt to change the biases themselves (for example, changing the way we automatically react to minorities). Other strategies focus on leaving the biases intact, but recognizing them and changing their effect on behavior.

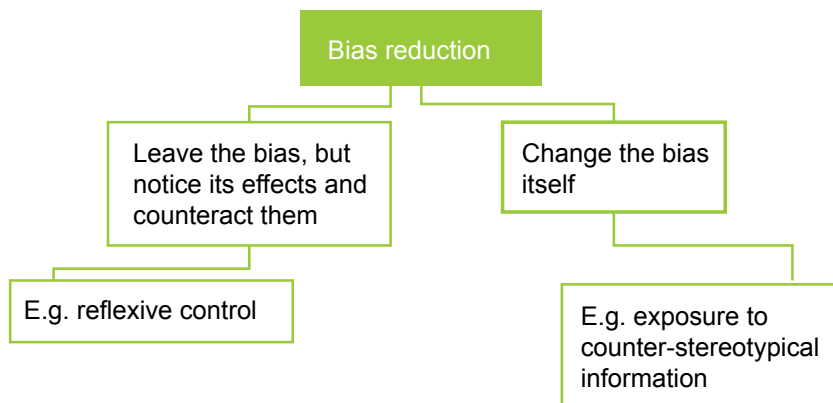
For example, one of the ways that proved to be effective in reducing stereotypes and prejudice is exposure to **counter-stereotypical information**. This can be something like watching films or simply picturing members of stereotyped groups engaging in counter-stereotypical behavior (female scientists, young presidents, sober rock stars, etc). In one study, Columb and Plant (2010) discovered the “Obama effect”: showing people a picture of Barack Obama or even simply his name resulted in a temporary reduction of stereotypes and prejudice against black people. This strategy attempts to change the bias itself.

Another approach is to leave the biases intact (let them be), but learn to notice their effects and counteract these effects when they become undesirable.

This is a hot topic of research. There are some findings that suggest that training yourself to actively counteract the biases that you are aware of may have positive results. Coming back to the shooter bias paradigm, in the work of Mendoza et al. (2010) this strategy is called “**reflexive control**”. Before starting the task, participants in their studies were instructed to use one of the rules:

- If I see a person, then I will ignore his face!
- If I see a person with a gun, then I will shoot!
- If I see a person with an object, then I will not shoot!

As you can see, all of these rules are aimed at separating the relevant aspect of the situation (gun versus a no-gun object) from the irrelevant aspect (race). Before the start of the experiments, participants were instructed to repeat the rule three times and write it down. Results showed that racial bias indeed decreased.



This is good news! It means that we can design relatively simple strategies that will prevent our implicit biases from acting in negative ways. Note that with such strategies, the bias itself is not targeted. We let the bias be, we just try to control the consequences.

This is a little like installing anti-virus software on your system riddled with viruses. It involves work and self-discipline, but it may prove to be fruitful.

KEY IDEA: Bias reduction is possible to some extent. We can either try to change the bias itself or mitigate its effects on behavior.

Critical thinking extension

We may agree at this point that bias in personal knowledge is controllable to some extent, but controlling it is a difficult task that requires constant cognitive effort and perhaps years of specially focused training.

At the heart of this cognitive effort lies the concept of **reflexivity**. This concept comes from the human sciences. It means the process of considering how the researcher's own mental processes may have influenced results of the research. For example, when an anthropologist observes a remote primitive tribe and sees that they engage in a lot of aggressive behavior, she may conclude that "the tribe in general seems very aggressive to me, but then again when I started this observation I expected them to be a violent tribe. I could have a tendency to notice aggressive behavior and overlook acts of kindness, so my observations should be corroborated by another researcher who does not have such background expectations". This is an example of reflexivity in social research – being aware of a possibility of biased judgment and taking it into account when presenting results.

Where else is reflexivity important? To what extent do you think it is important in areas of knowledge such as history and mathematics?

To what extent can bias in research be reduced through researcher reflexivity? How is it different in different areas of knowledge? (#Perspectives)

If you are interested...

Read the article "Debiasing: How to reduce cognitive biases in yourself and others" on the website *Effectiviology*.

To what extent do you think debiasing strategies are effective?

Take-away messages

Lesson 9. A what-if thought experiment shows that a bias-free society is not a desirable situation. In any case, bias elimination does not seem to be a possibility, but bias reduction could be possible to some extent. Research shows that bias reduction may take one of two forms: either changing the bias itself or leaving it intact but changing its effect on behavior. A large role in bias reduction is played by counter-stereotypical information. Exposing oneself intentionally to counter-evidence may be beneficial. Reflexive control is another key strategy of bias reduction. In this strategy, reflexivity is used to recognize the bias and consciously counteract its effects on thinking and behavior. Bias reduction involves a lot of work and self-discipline.

Lesson 10 - Compos mentis

Learning outcomes

- [Knowledge and comprehension] What is *compos mentis*?
- [Understanding and application] What are some arguments for and against moral responsibility for the outcomes of implicit, uncontrollable biases in personal knowledge?
- [Thinking in the abstract] To what extent are we morally responsible for the outcomes of biases that we are not aware of and can't control?

Recap and plan

We have established in the previous lessons that human mental software is riddled with biases. Many of these biases seem unavoidable, even when you are aware of them. Many biases are implicit – this means that they affect our thinking while we are certain that they don't. All this implies one thing: although it may feel like I am in control of my own mind, I am really not.

This has an important ethical dimension to it. The question is, since my biases are implicit and unavoidable, should I be held responsible for them?

Since my biases are implicit and unavoidable, should I be held responsible for them?

This is a question with numerous ethical and legal implications. When someone commits a crime, the judge takes into account whether or not the crime was intentional and whether or not it was **compos mentis**. *Compos mentis* is a Latin expression meaning “having full control of one's mind”. If the criminal did not control his actions at the time of committing the crime (in other words, the criminal was **non compos mentis**), we send him to a mental hospital rather than jail. We do not hold him accountable for his actions and we want him to get treatment rather than punishment.

Can this be extrapolated to other situations? When a stock broker loses millions of dollars' worth of assets because they made a decision that was way too risky, was that their fault or was it the mental software that failed them? When a pilot misreads some data on the dials and puts passengers' lives in danger, is the pilot morally responsible for the mistake or was that the natural limitations of human perception and cognition that we need to blame?

There are several approaches to answer these ethical questions – the argument from awareness, the argument from control and the deep self argument.

Key concepts

Compos mentis / non compos mentis, argument from awareness, argument from control, deep self argument

Other concepts used

Awareness, moral responsibility

Themes and areas of knowledge

Theme: Knowledge and the knower

Should knowers be held responsible for their implicit biases? (#Ethics)

Does culpability of an action or a decision depend on the person's amount of self-control? (#Ethics)



Image 31. Most of the time our brain is on autopilot

The argument from awareness

The argument from awareness states that biases are blameworthy only when the subject is consciously aware of them. If you do not suspect you have a bias, then you cannot be blamed for it. Someone who lives in a very sexist society, for example, will probably exhibit sexist attitudes and behavior without ever realizing that there is something wrong with those behaviors. According to the argument from awareness, such sexist attitudes are not blameworthy.

Is bias blameworthy if the knower is not aware of it?
(#Ethics)

The problem is that, if we follow the argument from awareness strictly, we must admit that lack of education is a sufficient excuse for immoral actions. Well-educated people will be more aware of their biases, so they hold more moral responsibility for their actions. It makes some sense, but the flipside doesn't: uneducated people are less morally responsible for their immoral behavior. Would you agree, for example, that uneducated people are less morally responsible for racism or sexism?

One way to stick to the argument from awareness and at the same time avoid this problem is to say that it doesn't matter what people *are* aware of, what matters is what they *ought to be* aware of. We may be held accountable for biases that we do not know if we *can* potentially know them. For example, if I am an uneducated person holding racist beliefs, I am still morally accountable for these beliefs as long as I have access to educational resources that I can use if I want. This is somewhat like driving a car without a license: if you get into an accident, you cannot just excuse yourself by saying "Oh, I am not to blame for this accident because I don't have a license and I can't drive". The point is, you *could* get a license and learn to drive, but you didn't.

From this point of view:

- A judge is held accountable for biases in judgment because it is their job to make judgments as impartial as possible, so they ought to take every effort to become aware of their implicit biases. They ought to read available professional literature, carefully consider alternative opinions, reflect on how and why they make decisions.
- A judge from 2020 is more morally responsible for implicit biases in their decisions than a judge from 1960. Back in 1960, scientific research on implicit cognitive biases was in its early stages; humanity was just beginning to get a grip of the idea that our mental software is full of bugs. It has all changed now. Since this knowledge is publicly available, the judge ought to have it, especially if they are involved in high-stakes decision making. They are more morally responsible now for their cognitive biases than they would have been 60 years ago.

Are humans becoming more morally responsible for their biases over the course of time?
(#Ethics)

The argument from control

The argument from control holds that we can only be morally responsible for actions that are within our control. Even if we are aware of a bias, we should not be morally responsible for its effects if we are not in the power to control it.

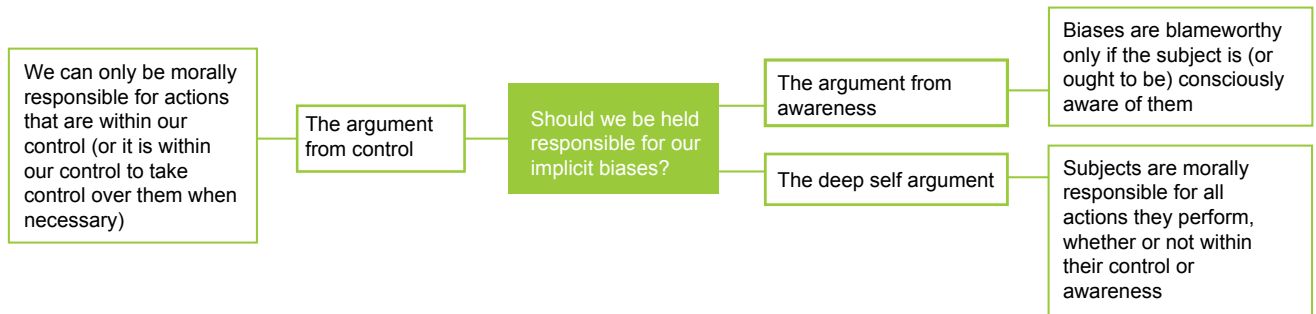
Am I morally responsible for not being able to save a friend from drowning if I tried, but my body was not strong enough to swim against a current? Probably not. I am only blameworthy, it seems, if there's a choice between A and B and it is within my willpower to choose either. This also seems to apply to implicit cognitive biases. Since many of them reside within System 1, they are largely automatic and unconscious. When my brain is on autopilot, I am not really controlling it. Yes, I can override autopilot when necessary (and then I probably become morally responsible for what happens), but most of the time I must rely on autopilot because my cognitive resources are so limited.

However, one might argue that there exists a degree of moral responsibility even when automatic and relatively unconscious actions are involved. Our responsibility may lie not with

the autopilot itself, but with knowing when to override it. Just like in a real airplane, the pilot cannot simply turn on the automatics, sit back, relax and blame whatever happens on the machine! The pilot is trained to recognize when it is better to rely on autopilot and when it is time to take over. A failure to recognize the crucial moment may well be within the pilot's moral responsibility.

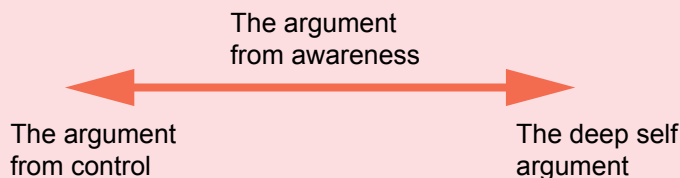
The deep self argument

The **deep self argument** claims that subjects can be held morally accountable for all actions they perform, whether or not those actions are within their conscious control or awareness. In other words, even if an action is performed by a part of me that I am not aware of or not in control of, it is still part of my "deep self". I find this position a little scary (am I alone?) – it means that I am morally responsible for all the actions of the horse I am trying to ride, even though this horse has a mind of its own.



Critical thinking extension

The three arguments presented here form a kind of a continuum. On one extreme, the deep self argument claims that people should be held responsible for all actions and their consequences, even if these actions were a result of deeply implicit biases that the person was not aware of and was not able to control. On the other extreme, the argument from control claims that even if we are aware of a bias, we should not be morally responsible for its outcomes if we cannot control them. The argument from awareness takes the middle ground, claiming that if we are aware of a bias (or can potentially become aware of it), then the responsibility for the outcomes lies with us.



Do you think these ethical considerations are especially applicable to experts who are in a position to make high-stakes decisions affecting other people's lives and well-being? Examples include judges, surgeons, commercial airline pilots, military leaders and presidents.

Even on a much smaller scale, and in everyday thinking and decision-making, do you think we hold moral responsibility for our biased perceptions, attitudes, opinions and utterances?

Interestingly, the more educated and self-aware you become, the less the *non compos mentis* excuse applies to you. Indeed, education is a curse, and greater knowledge implies greater responsibility.

Who should be held accountable for negative consequences of implicit bias? (#Ethics)

If you are interested...

Study the article “Understanding the law: culpable mental states” (June 26, 2018) on the *U.S. & Texas Lawshield Blog*. This gives you an idea of how the problem of culpability is tackled in today’s law.



Read and watch Willingham and Marco’s publication “She took her life, but he’s accused of helping her and filming it. Is it murder?” (October 21, 2017) for *CNN*. It is a story about a teen who was charged with his friend’s suicide. This raises some questions about the limits of criminal culpability.



Take-away messages

Lesson 10. The fact that our mental software is riddled with biases (many of which are implicit and beyond our conscious control) raises an ethical question: if we cannot control biases in our personal knowledge, to what extent should we be held morally responsible for outcomes of such biases? The concept of *compos mentis* (“having full control of one’s mind”) applies here. Although the concept is widely used in legal practice, it is currently limited to psychiatric cases. However, the problem is philosophical – to what extent are even mentally healthy people in full control of their mind? There are three main approaches to answering this ethical question. The argument from control states that we should not be held morally responsible for outcomes of biases that we are aware of, but cannot control. The argument from awareness states that we should not be held accountable for outcomes of biases that we are not aware of. The deep self argument assumes moral responsibility for all biases, both controllable and not.

Back to the exhibition

I am looking once again at my map of turbulence. I can safely say that I am a lot more confused than I was 10 lessons ago. I am not confused by the map, but by what it represents in terms of personal bias. Well, maybe not confused – more like I can see more sides to it.

Even before the journey that we undertook in this unit, I had realized that some of my fears were irrational. Now, I understand that some of my beliefs systematically deviate from shared knowledge because of distortions introduced by this irrational fear. I overestimate the danger of turbulence because I do not feel comfortable experiencing it, and I underestimate the danger of cars because I am so used to car travel.

In addition to that, I now wonder where these biases come from. They must be connected to my personal and cultural experiences. It is true, though I have travelled by air quite frequently, I never really experienced turbulence that could be categorized as severe. Perhaps it is lack of personal experience that causes me to fill the gaps with assumptions. I also wonder if fear of turbulence is a meme. These turbulence maps are pretty popular, so it must be a meme. Footage of severe turbulence quickly becomes viral, so this meme successfully replicates itself in people's minds. I wonder what the evolutionary advantage of this meme could be. Why is it there?

If it is indeed a meme that is at work here, then perhaps I am nothing more than a vessel meant to run an experimental simulation. Mother Nature infected me with this meme to see how it plays out. OK, I guess I am glad to participate in this global simulation and contribute some data to the cause.

I also wonder if it is even possible for me to reduce bias or even completely eliminate it. Overestimating the danger of turbulence really is just a tip of the iceberg. This is a bias I consciously recognize in myself. Behind it there is a whole army of biases that I am not even aware of, many of them – I am pretty sure – much worse than this one. Can I ever bring them to light and “debias” myself?

I have seen that it is possible to some extent, but also that it requires purposeful and consistent effort, struggling with my own self and slowly trying to gain control over my own mind.

Once I realized that, to what extent am I morally responsible to actually follow this path? It is not easy, so I might prefer to simply stay oblivious to the bugs in my mental software. On the other hand, if it is true that greater knowledge implies greater responsibility, I must fight them now that I realized they are there.

I sigh. It all started with an innocent map of turbulence. Ten lessons later it turned into a set of questions that make me rethink my whole existence. Perhaps turbulence is not what I should be afraid of, after all. Perhaps I should be afraid of my own self.